

KLASIFIKASI PRODUK BERDASARKAN JUMLAH TRANSAKSI PEMBAYARAN DI EDU POS SMKN 2 PADANG

Mutia Nurdina Wati¹, Yomei Hendra²

^{1,2}Program Studi Informatika, Adzkie Sumatera Barat, Indonesia

Article Info

Article history:

Received, 25 Juni 2024
Revised, 20 Juli 2024
Accepted, 01 Agustus 2024

Keywords:

Classification (Klasifikasi),
K-Nearest Neighbor,
Python,
Rapidminer,
Data Mining,
Transaksi Pembayaran.

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan produk berdasarkan jumlah transaksi pembayaran di Edu Pos SMKN 2 Padang. Edu Pos adalah layanan yang memfasilitasi pembayaran berbagai tagihan, dan penelitian ini menggunakan metode K-Nearest Neighbor (KNN) untuk mengelompokkan pelanggan menjadi pelanggan rutin dan sporadis berdasarkan data transaksi selama periode Januari hingga Desember 2023. Proses data mining meliputi tahapan data selection, preprocessing, transformasi data, dan normalisasi. Hasil analisis menunjukkan bahwa metode KNN berhasil mengklasifikasikan data dengan akurasi sebesar 90,28%. Hasil klasifikasi menunjukkan bahwa produk yang paling sering dibayar adalah PDAM Kota Padang dan PLN Postpaid, sementara produk yang jarang dibayar adalah pengisian ShopeePay, GoPay, dan pulsa. Temuan ini memberikan wawasan baru untuk meningkatkan layanan dan pengambilan keputusan di Edu Pos SMKN 2 Padang. Penelitian selanjutnya disarankan untuk menggabungkan dua metode algoritma data mining untuk hasil yang lebih variatif dan bernilai informasi tinggi.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

Mutia Nurdina Wati,
Program Studi Informatika ,
Universitas Adzkie Sumatera Barat,
Jl. Taratak Paneh No. 7 Korong Gadang, Kalumbuk, Kec. Kuranji, Kota Padang, Sumatera Barat.
Email: mutianurdinawati02@gmail.com

1. PENDAHULUAN

Edu Pos merupakan layanan yang disediakan oleh SMKN 2 Padang untuk memfasilitasi pembayaran berbagai biaya tagihan, seperti tagihan BPJS, listrik, asuransi, dan biaya lainnya. Edu Pos melakukan kerja sama dengan Pos Finansial Indonesia, dan Edu Pos merupakan salah satu agen dari Pos Finansial Indonesia. Layanan ini bertujuan untuk memberikan kemudahan bagi masyarakat di sekitar untuk melakukan pembayaran dan kebutuhan tagihan lainnya. Dengan adanya Edu Pos, diharapkan proses pembayaran dapat dilakukan dengan lebih efisien dan teratur. Seiring dengan meningkatnya jumlah pelanggan yang menggunakan layanan Edu Pos, pihak sekolah membutuhkan informasi yang lebih mendalam tentang Klasifikasi pelanggan berdasarkan jumlah transaksi pembayaran berarti pelanggan akan dikelompokkan atau dikategorikan berdasarkan seberapa sering atau banyak mereka melakukan transaksi pembayaran melalui Edu Pos SMKN 2 Padang.

Salah satu pendekatan yang dapat digunakan untuk memperoleh informasi tersebut adalah dengan menggunakan metode K-Nearest Neighbor (KNN) untuk melakukan klasifikasi pelanggan

berdasarkan jumlah transaksi pembayaran yang mereka lakukan. Metode KNN adalah salah satu metode klasifikasi dalam data mining yang berbasis pada perhitungan jarak antara data yang akan diklasifikasikan dengan data pelatihan yang ada. Dalam konteks ini, data pelatihan dapat berupa informasi jumlah transaksi pembayaran dari pelanggan yang telah diklasifikasikan sebelumnya. Klasifikasi dapat dilakukan dengan membagi pelanggan menjadi dua kategori, yaitu pelanggan yang rutin dan pelanggan sporadis (jarang). Untuk pengolahan data transaksi dilakukan selama periode 1 tahun yaitu dari bulan Januari s/d Desember 2023.

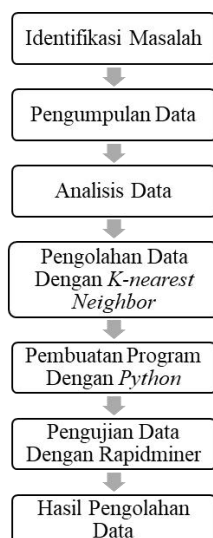
Dalam penelitian ini, pengolahan data dan implementasi model klasifikasi dilakukan menggunakan bahasa pemrograman Python dan perangkat lunak RapidMiner. Python dipilih karena memiliki berbagai pustaka yang kuat dan fleksibel untuk analisis data dan pembelajaran mesin, seperti pandas untuk manipulasi data, scikit-learn untuk algoritma pembelajaran mesin, dan matplotlib serta seaborn untuk visualisasi data. Proses pengolahan data dalam Python mencakup tahap-tahap seperti pembersihan data (data cleaning), eksplorasi data (exploratory data analysis), dan persiapan data (data preprocessing). RapidMiner, di sisi lain, adalah perangkat lunak yang menyediakan platform terpadu untuk proses data mining dan pembelajaran mesin. RapidMiner menawarkan antarmuka pengguna yang intuitif dengan dukungan drag-and-drop, sehingga memudahkan dalam membangun, menguji, dan mengimplementasikan model pembelajaran mesin. RapidMiner juga dilengkapi dengan berbagai operator untuk berbagai tugas data mining seperti pemrosesan data, transformasi data, dan evaluasi model. Dalam penelitian ini, RapidMiner digunakan untuk membangun dan mengevaluasi model K-Nearest Neighbor (KNN) yang digunakan untuk klasifikasi produk berdasarkan jumlah transaksi pembayaran.

Dengan menggabungkan Python dan RapidMiner, penelitian ini memanfaatkan keunggulan masing-masing alat untuk menghasilkan analisis yang mendalam dan model klasifikasi yang akurat. Python digunakan untuk tahapan awal pengolahan dan eksplorasi data, sementara RapidMiner digunakan untuk membangun dan mengevaluasi model klasifikasi secara efisien. Dalam penelitian yang dilakukan oleh A. Tangkelayuk et al. (2022) [1], Algoritma KNN merupakan suatu metode untuk mengklasifikasi objek dari sebuah datauji berdasarkan data latih yang memiliki jarak paling dekat dengan objek tersebut. Penelitian lain G. Dwidasmara et al (2021) [2] K-Nearest Neighbor (KNN) adalah salah satu metode pengelompokan dalam data mining, dimana KNN mengelompokkan sekumpulan data berdasarkan data pembelajaran yang telah diklasifikasi atau diberi label. KNN masuk kedalam kelompok supervised learning yaitu hasil query yang baru dikelompokkan instans berdasarkan mayoritas kedekatan dengan kategori yang ada di KNN.

Dalam konteks penelitian ini di SMKN 2 Padang, penggunaan metode KNN untuk klasifikasi pelanggan Edu Pos berdasarkan jumlah transaksi pembayaran mereka memiliki potensi yang menjanjikan. Meskipun belum ada penelitian yang secara khusus mengimplementasikan metode KNN dalam konteks Edu Pos di SMKN 2 Padang, hasil-hasil penelitian yang dilakukan pada industri dan layanan serupa telah menghasilkan wawasan yang berharga. Oleh karena itu, penelitian ini bertujuan untuk menerapkan metode KNN dalam klasifikasi pelanggan Edu Pos SMKN 2 Padang berdasarkan jumlah transaksi pembayaran mereka. Diharapkan hasil penelitian ini dapat memberikan kontribusi positif dalam peningkatan layanan Edu Pos dan pengambilan keputusan yang lebih baik oleh pihak sekolah.

2. METODOLOGI

Penelitian ini menggunakan metode K-Nearest Neighbor (KNN) untuk melakukan klasifikasi produk berdasarkan jumlah transaksi pembayaran yang mereka lakukan. Data yang digunakan adalah data transaksi pembayaran selama periode Januari hingga Desember 2023 di Edu Pos SMKN 2 Padang. Tahapan yang harus dilakukan oleh peneliti sebelum melakukan penelitian adalah desain penelitian. Desain penelitian merupakan tahapan atau proses yang digunakan untuk merencanakan suatu penelitian. Desain penelitian bertujuan untuk mempermudah peneliti dalam melakukan penelitian. Berikut ini adalah langkah-langkah dalam desain penelitian:



Gambar 1. Tahapan Penelitian

Keterangan dari tahapan pada penelitian ini sebagai berikut :

1. Identifikasi Masalah : Seiring meningkatnya jumlah pelanggan Edu Pos SMKN 2 Padang, pihak sekolah membutuhkan informasi mendalam tentang karakteristik dan perilaku pelanggan dalam melakukan transaksi pembayaran. Namun, belum ada metode khusus yang diterapkan untuk mengklasifikasikan pelanggan berdasarkan jumlah transaksi pembayaran mereka. Oleh karena itu, penelitian ini bertujuan untuk menerapkan metode K-Nearest Neighbor (KNN) dalam mengklasifikasikan pelanggan Edu Pos berdasarkan jumlah transaksi pembayaran.
2. Pengumpulan Data : Pengumpulan Data Transaksi Pembayaran Pada tahap ini, data transaksi pembayaran pelanggan Edu Pos SMKN 2 Padang selama periode 1 tahun (Januari - Desember 2023) akan dikumpulkan. Data yang dikumpulkan meliputi informasi seperti tanggal transaksi, jenis pembayaran, jumlah nominal, dan identitas pelanggan.
3. Analisis Data : Pada tahapan ini melakukan pembersihan data transaksi pembayaran dari data yang tidak penting, Melakukan transformasi data ke dalam format yang sesuai untuk digunakan dalam *K-Nearest Neighbor* (KNN). Membagi data transaksi pembayaran ke dalam data pelatihan (training set) dan data pengujian (testing set).
4. Pengolahan data dengan *K-Nearest Neighbor* : Pengolahan data dan data yang telah diperoleh dari analisa dan diolah kedalam metode *K-Nearest Neighbor* untuk melakukan klasifikasi data pada kelas tertentu.
5. Pembuatan Program Menggunakan *Python* : Pada tahapan ini dengan membuat program untuk mengolah data menggunakan python kedalam metode *K-Nearest Neighbor* untuk melakukan klasifikasi data pada kelas tertentu.
6. Pengujian data dengan Rapidminer : Setelah data diolah menggunakan metode *K-Nearest Neighbor* kemudian data diuji menggunakan *software* Rapidminer untuk mendapatkan hasil yang selaras dengan perhitungan secara manual.
7. Hasil Pengolahan Data : Setelah melakukan pengujian menggunakan pembuatan program dengan python dan menggunakan *software* Rapidminer maka akan mendapatkan hasil dan ini merupakan akhir dari proses penelitian yang dilakukan oleh peneliti.

Metode Pengumpulan Data

Penelitian ini menggunakan 2 teknik pengumpulan data yaitu sebagai berikut:

1. Wawancara merupakan suatu cara mengumpulkan data untuk memperoleh informasi langsung dari sumbernya. di mana peneliti memperoleh informasi langsung dari ibu Nurma selaku pengelola edu pos SMKN 2 Padang.

2. Studi Pustaka, Peneliti melakukan studi pustaka untuk mendapatkan informasi secara langsung dari beberapa sumber yakni buku, jurnal, web yang berkaitan langsung dengan objek penelitian.

Pengolahan Data Mining

K-Nearest Neighbor (K-NN)

Berdasarkan tahapan data mining untuk algoritma K-Nearest Neighbor berikut ini adalah langkah-langkah algoritma K-NN:

1. Menentukan parameter nilai k (di mana nilai k pada penelitian ini menggunakan nilai 3). untuk menghitung antara jarak data testing dan data training.
2. Mengurutkan data training dengan berdasarkan nilai k (3) dari yang terkecil. untuk menetapkan kelas, berdasarkan nilai k (3) dengan jumlah terbanyak.

Data Selection

Pengumpulan data transaksi pembayaran pelanggan Edu Pos SMKN 2 Padang selama periode 1 tahun (Januari - Desember 2023). Data yang dikumpulkan meliputi informasi seperti layanan, produk, nominal, saldo sebelum, saldo sesudah, dan waktu. Jadi pada tahap data selection, peneliti mengumpulkan data transaksi pembayaran pelanggan Edu Pos SMKN 2 Padang selama satu tahun penuh yaitu dari Januari hingga Desember 2023 yang mencakup detail seperti tanggal transaksi, jenis pembayaran yang dilakukan, jumlah nominal yang dibayarkan, serta identitas pelanggan yang melakukan transaksi tersebut.

Preprocessing

1. Data Cleaning (Pembersihan Data)
Menghapus data transaksi yang tidak relevan atau di luar periode Januari - Desember 2023, memeriksa dan menangani data transaksi yang mungkin memiliki nilai yang hilang (missing value) pada kolom tanggal transaksi, jenis pembayaran, jumlah nominal, atau identitas pelanggan
2. Data Transformation (Transformasi Data)
Mengubah format tanggal transaksi menjadi format yang konsisten, misalnya yyyy-mm-dd, melakukan encoding pada jenis pembayaran menjadi nilai numerik, misalnya: 1 = BPJS, 2 = Listrik, 3 = Asuransi dan seterusnya, tidak perlu normalisasi pada jumlah nominal, karena nilai tersebut sudah numerik
3. Data Partitioning (Partisi Data)
Membagi dataset transaksi menjadi data training dan data testing dengan rasio tertentu, misalnya 80% data training dan 20% data testing, pembagian data dapat dilakukan dengan teknik simple random sampling

Setelah melalui tahapan preprocessing di atas, data transaksi pembayaran pelanggan Edu Pos SMKN 2 Padang sudah siap untuk digunakan dalam metode K-Nearest Neighbor untuk klasifikasi pelanggan berdasarkan jumlah transaksi pembayaran.

Transformation

Tahapan Transformation merupakan tahap merubah data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses transformasi dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Pada tahapan ini dari seluruh data operasional didapatkan data pengelompokan atribut yang digunakan untuk proses transformasi data mining, yaitu atribut bulan dan klasifikasi sebagai kriteria data yang menjadi target dalam proses mining.

Data Mining

Tahap ini merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu berdasarkan proses KDD secara keseluruhan. Metode

yang digunakan pada penelitian ini adalah metode K-Nearest Neighbor (KNN) dimana metode ini memiliki atribut yang diinisialisasikan sebagai k, yaitu jumlah tetangga yang dijadikan acuan pada KNN, nilai k adalah bilangan bulat positif, berjumlah kecil dan ganjil.

Adapun tahapan pengerjaan metode KNN dalam penelitian ini adalah sebagai berikut :

1. Penentuan nilai k. Penentuan nilai k yang digunakan dalam klasifikasi tidak memiliki aturan yang baku, namun pada penelitian ini nilai k yang digunakan adalah 3, 5, 7, 9, 11, 13, 15.
2. Perhitungan jarak antar data training dan data uji(test). Teknik perhitungan jarak yang digunakan dalam metode KNN ini adalah Jarak Euclidean

Distance. Dengan rumus sebagai berikut :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Contoh perhitungan jarak dalam KNN menggunakan Euclidean Distance akan dijelaskan pada bab selanjutnya dengan menggunakan data real.

3. Pengurutan data hasil perhitungan. Jarak yang telah didapatkan kemudian diurutkan dari yang paling dekat jaraknya sampai yang paling jauh (ascending).
4. Menentukan kelompok data hasil uji berdasarkan label mayoritas dari k tetangga terdekat.

Evaluasi

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Pada tahap ini didapatkan pola penjualan elektronik dari proses data mining dengan metode K-Nearest Neighbor, pola atau informasi yang dihasilkan dari proses data mining adalah berupa rules yang didapat dari perhitungan.K-Nearest Neighbor.

3. HASIL DAN PEMBAHASAN

3.1. Data Selection

Data yang digunakan dalam penelitian ini yaitu data transaksi pembayaran berdasarkan pembayaran 1 tahun terakhir dari 2023 yang berasal dari Edu pos SMKN 2 Padang. Kemudian data tersebut diseleksi dan akan digunakan untuk diolah dalam penentuan karakteristik atau pola pelanggan yang melakukan transaksi berdasarkan banyak transaksi dan jumlah pembayaran. Adapun atribut yang digunakan dalam penentuan karakteristik pelanggan adalah atribut tanggal transaksi, jenis pembayaran, jumlah nominal, serta identitas pelanggan yang melakukan transaksi tersebut.

Tabel 1. *Data Selection*

No	Produk	Nominal	Waktu
1	BPJS	Rp 317.500	2 Januari 2023
2	Home Credit Indonesia	Rp 665.800	2 Januari 2023
3	Adira Finance	Rp 740.000	4 Januari 2023
4	Adira Finance	Rp 669.000	4 Januari 2023
5	Adira Finance	Rp 260.000	4 Januari 2023
6	Adira Finance	Rp 912.500	4 Januari 2023
7	Home Cedit Indonesia	Rp 801.900	5 Januari 2023
8	Home Cedit Indonesia	Rp 388.000	6 Januari 2023
9	Adira Finance	Rp 358.500	6 Januari 2023
10	Home Cedit Indonesia	Rp 460.400	9 Januari 2023
...
1121	Adira Finance	Rp 1.307.500	22 Desember 2023

3.2. Preprocessing

Preprocessing adalah langkah-langkah awal yang dilakukan untuk mempersiapkan data mentah sebelum digunakan dalam analisis lebih lanjut atau model machine learning. Tujuan utama dari preprocessing adalah untuk membersihkan dan mengubah data mentah sehingga menjadi lebih mudah dianalisis dan lebih bermanfaat untuk model yang akan digunakan. Mengganti type data dimana “PRODUK” type data Object, “NOMINAL” type data Int, “TANGGAL” type data Int, “BULAN” type Int dimana “BULAN”.

#	Column	Non-Null Count	Dtype
0	PRODUK	1122 non-null	object
1	NOMINAL	1122 non-null	int64
2	TANGGAL	1122 non-null	int64
3	BULAN	1122 non-null	int64

dtypes: int64(3), object(1)

Gambar 2. Tipe Data

Saat preprocessing data tidak terdapatnya data kosong dan duplikat data di dalam data mentah.

	PRODUK	NOMINAL	TANGGAL	BULAN
0	BPJS KESEHATAN VA KELUARGA (1 bulan bayar)	317500	2	1
1	Home Credit Indonesia	665800	2	1
2	ADIRA FINANCE	740000	4	1
3	ADIRA FINANCE	669000	4	1
4	ADIRA FINANCE	260000	4	1
...
1116	PDAM KOTA PADANG	139822	20	12
1117	PLN POSTPAID	307750	20	12
1118	PLN POSTPAID	64300	20	12
1119	ADIRA FINANCE	56539	20	12
1120	PLN POSTPAID	221618	20	12

Gambar 3. Data Setelah Outlier

Setelah melakukan outlier terdapat 1 data yang kosong pada data mentah maka data yang didapatkan setelah melakukan preprocessing yaitu 1120 data.

3.3 Transformasi

Pada tahap transformation ini hasil dari perubahan type “PRODUK” diubah dari teks menjadi angka. Transformasi ini dilakukan melalui encoding, dimana setiap produk diberikan kode unik berupa angka. Angka lebih efisien untuk disimpan dan diproses oleh komputer dibandingkan dengan teks. Banyak algoritma machine learning membutuhkan input data dalam format numerik.

	PRODUK	NOMINAL	TANGGAL	BULAN
0	4	317500	2	1
1	13	665800	2	1
2	2	740000	4	1
3	2	669000	4	1
4	2	260000	4	1
...
1116	22	139822	20	12
1117	25	307750	20	12
1118	25	64300	20	12
1119	2	56539	20	12
1120	25	221618	20	12

Gambar 4. Data Transformasi

3.4 Normalisasi

Normalisasi adalah proses mengubah nilai-nilai numerik dalam data agar berada dalam skala yang sama, biasanya untuk memperbaiki kinerja algoritma machine learning dan membuat analisis lebih mudah. Berikut ini tabel normalisasi

Tabel 2. Data Normalisasi

Produk	Nominal	Tanggal	Bulan
-1.313147	0.138092	-1.719620	-1.917715
-0.539328	1.414364	-1.719620	-1.917715
-1.485107	1.686254	-1.440106	-1.917715
-1.485107	1.426090	-1.440106	-1.917715
-1.485107	-0.072605	-1.440106	-1.917715
.....
0.234491	0.512972	0.796008	1.594625
0.492430	0.102365	0.796008	1.594625
0.492430	-0.789706	0.796008	1.594625
-1.485107	-0.818145	0.796008	1.594625

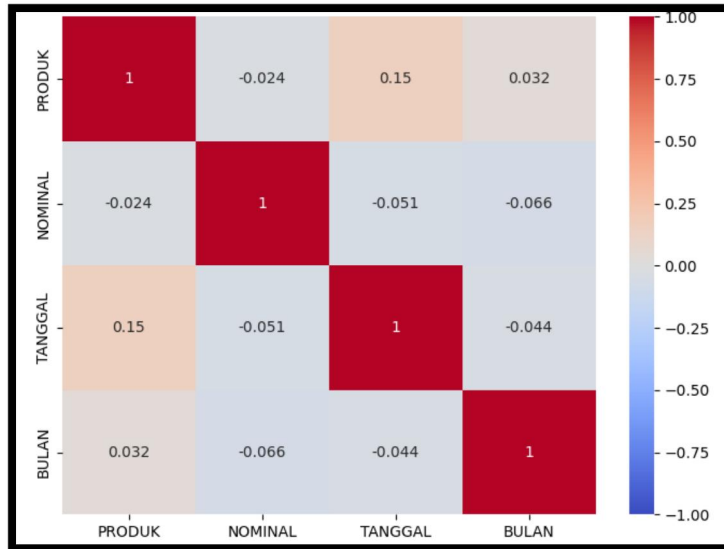
Tabel 3. Hasil Kolerasi

	Produk	Nominal	Tanggal	Bulan
Produk	1.000000	-0.023575	0.151023	0.031777
Nominal	-0.023575	1.000000	-0.051216	-0.065728
Tanggal	0.151023	-0.051216	1.000000	-0.044176
Bulan	0.031777	-0.065728	-0.044176	1.000000

Hasil analisis korelasi antara variabel-variabel dalam dataset ini menunjukkan bahwa sebagian besar hubungan antara variabel-variabel tersebut sangat lemah. Korelasi antara jenis produk dan nominal pembayaran adalah -0.023575, menunjukkan hampir tidak ada hubungan linier antara kedua variabel ini. Korelasi antara jenis produk dan tanggal pembayaran sedikit lebih tinggi pada 0.151023, namun tetap menunjukkan hubungan yang lemah. Begitu pula, korelasi antara jenis produk dan bulan pembayaran adalah 0.031777, yang juga menunjukkan hampir tidak ada hubungan linier.

Selanjutnya, hubungan antara nominal pembayaran dan tanggal pembayaran memiliki korelasi -0.051216, dan hubungan antara nominal pembayaran dan bulan pembayaran memiliki korelasi -0.065728. Kedua nilai ini menunjukkan hampir tidak ada hubungan linier antara nominal pembayaran dengan tanggal maupun bulan pembayaran. Korelasi antara tanggal pembayaran dan bulan pembayaran juga sangat lemah dengan nilai -0.044176.

Secara keseluruhan, hasil ini mengindikasikan bahwa tidak ada hubungan linier yang kuat antara variabel-variabel dalam dataset ini. Korelasi yang lemah ini menunjukkan bahwa perubahan dalam satu variabel tidak secara konsisten terkait dengan perubahan dalam variabel lain. Dengan demikian, jenis produk, nominal pembayaran, tanggal pembayaran, dan bulan pembayaran berfungsi secara independen dalam konteks dataset ini. Berikut ini merupakan gambar Heatmap Korelasi.

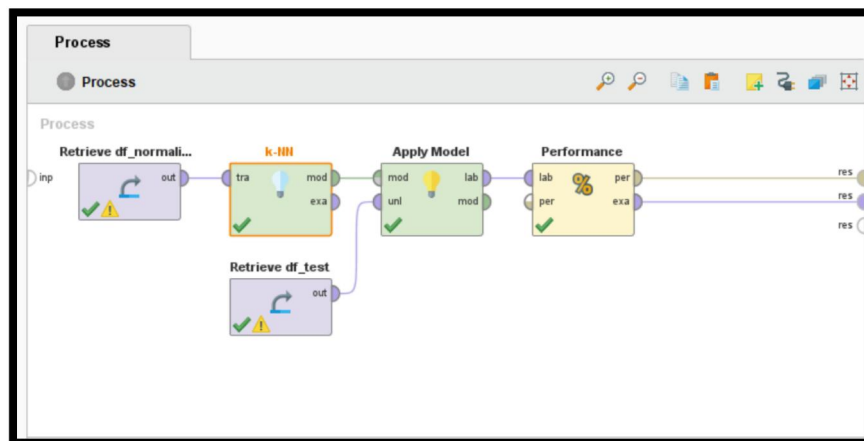


Gambar 5. Heatmap Korelasi

Dengan kata lain, data menunjukkan bahwa jenis produk, nominal pembayaran, tanggal pembayaran, dan bulan pembayaran sebagian besar independen satu sama lain dalam konteks dataset ini. Heatmap korelasi ini memberikan pandangan sekilas yang cepat dan jelas tentang hubungan linier antar variabel, membantu dalam memahami struktur data secara keseluruhan.

3.5 Implementasi Rapidminer

RapidMiner merupakan salah satu *software* data mining pengolahan data set untuk mencari pola data sesuai dengan tujuan dari pengolahan data tersebut, tidak semua algoritma yang ada dapat sesuai atau dapat mengolah data set yang ada, harus dilakukan penyesuaian pola data dan sesuai dengan tujuan dari pengolahan data tersebut.



Gambar 6. Model Sub Proses Clasification KNN

Criterion: accuracy

Table View Plot View

accuracy: 90.28%

	true PDA...	true PLN...	true TEL...	true BPJ...	true BPJ...	true BPJ...	true BUS...	true ADI...	true Mar...
pred. PD...	46	1	0	0	0	0	0	0	0
pred. PL...	2	45	0	0	0	0	0	0	0
pred. TE...	1	1	11	0	0	0	0	0	0
pred. BP...	0	0	0	1	0	0	0	0	0
pred. BP...	1	2	2	0	25	0	0	0	0
pred. BP...	0	0	0	0	0	1	0	0	0
pred. BU...	0	0	0	0	0	0	6	0	0

Gambar 7. Hasil Perhitungan Accuracy

Hasil Akurasi adalah ukuran seberapa sering model membuat prediksi yang benar. Dengan akurasi 90.28%, ini menunjukkan bahwa model secara keseluruhan memiliki kinerja yang baik.

```

KNNClassification
Weighted 3-Nearest Neighbour model for classification.
The model contains 1122 examples with 2 dimensions of the following classes:
BPJS KESEHATAN VA KELUARGA (1 bulan bayar)
Home Credit Indonesia
ADIRA FINANCE
TELKOM Jastel
TOYOTA ASTRA FINANCIAL SERVICES
FEDERAL INTERNATIONAL FINANCE (FIF)
Mandiri Utama Finance
PLN NON TAGIHAN
PULSA THREE 50K
PULSA THREE 30K
INDOMOBIL FINANCE INDONESIA
BUSSAN AUTO FINANCE
PLN POSTPAID
PDAM KOTA PADANG
BPJS KESEHATAN VA KELUARGA (2 bulan bayar)
    
```

Gambar 8. Model Klasifikasi KNN

Menggambarkan model pengelompokan produk berdasarkan pengklasifikasian pada metode *K-Nearest Neighbor* menggunakan tool *RapidMiner*.

3.6 Hasil Klasifikasi KNN

Tahap ini adalah tahap terakhir dari proses data mining rapidminer dimana setelah semua operator terhubung kemudian klik icon play tombol F11, maka akan muncul sebuah tab result, yang isinya sebuah prediksi dari seluruh data yang memenuhi nilai $k=3$, didapatkan hasil prediksi transaksi pembayaran seperti gambar berikut :

PerformanceVector

PerformanceVector:
accuracy: 90.28%

ConfusionMatrix:

True:	PDAM KOTA PADANG	PLN POSTPAID	TELKOM Jastel	BPJS KESEHATAN DENDA	BPJS KESEHATAN VA KELUA
PDAM KOTA PADANG:	46	1	0	0	0
PLN POSTPAID:	2	45	0	0	0
TELKOM Jastel:	1	1	11	0	0
BPJS KESEHATAN DENDA:	0	0	0	1	0
BPJS KESEHATAN VA KELUARGA (1 bulan bayar):	1	2	2	0	25
BPJS KESEHATAN VA KELUARGA (7 bulan bayar):	0	0	0	0	0
BUSSAN AUTO FINANCE:	0	0	0	0	6
ADIRA FINANCE:	0	0	1	0	0
Mandiri Utama Finance:	0	0	0	0	0
FEDERAL INTERNATIONAL FINANCE (FIF):	0	1	0	0	0
DANA 30K:	0	0	0	0	0
MEGA FINANCE:	0	0	0	0	0

Gambar 9. Hasil Performance Vector

Hasil dari memprediksi dengan benar kelas PDAM KOTA PADANG sebanyak 46 kali. Prediksi benar lainnya mencakup 45 untuk kelas PLN POSTPAID, 11 untuk TELKOM Jastel, 29 untuk BPJS KESEHATAN DENDA, dan 25 untuk BPJS KESEHATAN VA KELUARGA (1 bulan bayar). Jadi hasil dari klasifikasi produk, kategori “Rutin”dibayar adalah PDAM KOTA PADANG dan PLN POSTPAID. Sedangkan kategori “Sporadis” adalah pembelian SHOPPEPAY, GOPAY, dan PULSA.

4. KESIMPULAN

Penelitian ini berhasil menerapkan metode K-Nearest Neighbor (KNN) untuk mengklasifikasikan produk Edu Pos SMKN 2 Padang berdasarkan jumlah transaksi pembayaran mereka selama periode Januari-Desember 2023. Proses data mining dilakukan melalui beberapa tahap, meliputi data selection, preprocessing, transformasi data, dan normalisasi. Dari 1121 data awal, setelah preprocessing diperoleh 1120 data yang siap dianalisis.

Pada penelitian ini dilakukan pemodelan menggunakan *k- nearest neighbor* dengan menggunakan data yang diolah menggunakan *Python* untuk melakukan preprocessing dan EDA, untuk implementasi dan modeling menggunakan *Rapidminer* dengan hasil akurasi sebesar 90.28%. Dengan hasil klasifikasi produk yang rutin dibayar adalah PDAM KOTA PADANG sebanyak 46 kali dan PLN POSTPAID sebanyak 45 kali, untuk klasifikasi produk yang sporadis atau jarang dibayar yaitu pengisian SHOPPEPAY, GOPAY, dan PULSA. Penggunaan metode KNN dalam penelitian ini memberikan wawasan baru tentang produk apa yang paling sering dibayar di Edu Pos SMKN 2 Padang, yang dapat digunakan untuk meningkatkan layanan dan pengambilan keputusan di masa mendatang.

DAFTAR PUSTAKA

- [1] Afri, Y., Yosep, N., & Budi, P. (2021). KLASIFIKASI PENENTUAN CUSTOMER KREDIT MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *Jurnal Sains & Teknologi Fakultas Teknik*, 11(1), 64-71.
- [2] Amalia, Y. R. (2018). Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus: PT. Bintang Multi Sarana Palembang) (Doctoral dissertation, UIN RADEN FATAH PALEMBANG).
- [3] Halim, T. N., Martin, R., Hutasoit, R. K. B., & Aisyah, S. (2023). Klasifikasi Kepuasan Pelanggan Terhadap Platform E-Commerce dengan Metode K-Nearest Neighbor (K-NN). *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 8(2), 512-523.
- [4] Junaedi, E., Siregar, A., & Nurlaelasari, E. (2022). Implementasi C4. 5 Dan Algoritma K Nearest Neighbor Untuk Prediksi Kelayakan Pemberian Kredit Menggunakan RapidMiner Studio. *Scientific Student Journal for Information, Technology and Science*, 3(1), 83-90.
- [5] Mustaqim, K., Amaresti, F. A., & Dewi, I. N. (2024). Analisis Sentimen Ulasan Aplikasi PosPay untuk Meningkatkan Kepuasan Pengguna dengan Metode K-Nearest Neighbor (KNN). *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 11-20.
- [6] Susilo, A. (2023). Perbandingan Kinerja K-Nearest Neighbors dan Naive Bayes Untuk Klasifikasi Perilaku Nasabah Pada Pembayaran Kredit Bank. *Jurnal Sains dan Teknologi (JSIT)*, 3(2), 153-169.
- [7] Tangkelayuk, A. (2022). The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naive Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1109-1119.
- [8] Sumarlin. 2015. Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM. (<http://ejournal.undip.ac.id/index.php/jsinbis>) diakses 3 Maret 2018.